Proceedings of Science and Mathematics
Faculty of science, universiti teknologi malaysia
https://science.utm.my/procscimath
Volume 21 (2024) 157-168

Functional Data Analysis on Dengue Haemorrhagic Fever (DHF) in Malaysia

Mohamad Haziq Mohd Zam Zam, Muhammad Fauzee Hamdan*
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
*Corresponding author: mfauzee@utm.my

Abstract

Diseases or viruses is familiar things in our society and even become a gigantic phenomenon that can affect the total population of a country including dengue fever. This research aims to analyse the trend and pattern of Dengue Hemorrhagic Fever (DHF) cases in Malaysia through functional data analysis (FDA) technique based on past datasets. Data from Malaysia's Official Open Data Portal was utilized and two types of basis functions, B-spline and Fourier, were employed to create smooth curves. This study found that B-spline basis functions were most suitable for accurately depicting pattern and trend of DHF cases across 13 states and 1 federal territory over a 12-year period. By analyzing errors and curve smoothness, it was determined that a moderate number of basis functions which is around 10 were optimal for most states. This approach offers valuable insights for public health authorities to develop targeted prevention strategies and readiness measures for future DHF outbreaks in Malaysia.
Keywords: Dengue Hemorrhagic Fever; Functional Data Analysis; B-spline basis function; Fourier basis function.

Introduction

The number of dengue cases has increased all year round that encourage the growth of mosquitoes and their eggs, which in turn increases the population of Aedes mosquitoes that leads to the spreads of the dengue virus [1]. Dengue fever can develop into more severe forms like Dengue Hemorrhagic Fever (DHF), which can result in fatalities, bleeding, and a drastic reduction in blood pressure [2]. There is currently no particular treatment for dengue or DHF but there is the availability of appropriate medical care to decrease the fatality rate associated with severe dengue. Then, proper analyses should be carried out frequently as to observe the current situation related to dengue cases especially DHF case which can cause fatality in the community.

Thus, this project will provide better insight on DHF cases that has been occurred in Malaysia by using functional data analysis. Functional Data Analysis (FDA) was used with R programming language which will help to give detail analysis about cases of DHF in Malaysia through data obtained from Malaysia's Official Open Data Portal that involved cases in state Kedah, Kelantan, Negeri Sembilan, Pahang, Perak, Melaka, Johor, Perlis, Pulau Pinang, Sabah, Sarawak, Selangor, Terengganu and Wilayah Persekutuan Kuala Lumpur from 2010 to 2021. Therefore, detailed presentation of data involved DHF cases is produced in this research as an effort in order to reduce dengue case which can be done through observation of data patterns in all states in the country.

Literature Review

FDA in other field of studies

Functional data analysis (FDA) is a statistical technique are used to analyse the overall patterns, trends, and variability in any functional data, which are regarded as curves or functions. It is involved various approaches for modelling and analysing data that can be visualised as curves or functions. Analysis in FDA involved the linear and nonlinear methods, clustering and classification, and dimension reduction techniques such as functional principal component analysis [3]. FDA has been used by many researchers in various field of previous studies [4-9]. Shaadan [10] used FDA to estimate the missing values in data of air quality from air quality monitoring station at Petaling Jaya. The study used curve estimation technique through conversion of discrete data into curves with the implementation of several imputation methods. On the other hand, Ismail and Jamaludin [11] applied FDA to represent data of wind and rainfall through smoothing curves, variation of both climate parameters and analysing changes with determining the functional relationship among both data set. Oshinubi et al. [12] utilized FDA in analysing and modeling the pattern of the pandemic COVID-19 of different departments in French before and after vaccination through data obtained from public databases. It helped to process a large sample of epidemic data including longitudinal and cross-sectional data as to monitor and predict spread and impact of the pandemic.

Dengue case studies with various analysis approaches

Based on previous study, there are no specific study that apply FDA in dengue case study but various analysis was applied by researchers including spatiotemporal analysis [13], meta-analysis [14] and average nearest neighbour (ANN) analysis [15]. Furthermore, Chengathir et al. [16] utilized data mining techniques which specifically using R programming to analyse the dengue dataset and apply various classification algorithms to identify patterns and relationships within the data. Besides, Mohsin et al. [17] developed an accurate early detection system for dengue outbreaks in Malaysia using an artificial immune system. The prediction for future dengue outbreaks has been executed by Husin and Salim [18] in their study through machine learning technique. Neural network model (NNM) and non-linear regression model (NLRM) were used in this study together with time series data, location data, and rainfall data in their research.

Methodology

Descriptive statistic

Basic statistic was used to calculate the mean number of cases over 12-year period for each state by each week of 53 weeks. The mean, commonly referred to as the average which is crucial in producing a summary value that indicates the average or central value of a dataset. Then, these average cases will be used to smooth the curve for each state as to observe the trend and pattern of cases. Thus, mean was calculated by using the formula of $\bar{x} = \frac{\sum x}{n}$, where $\sum x$ is sum value of data which is the total cases for 12-year period for each week while n denotes the amount of data.

Functional Basis Smoothing

Data can be represented in curve or function form. Then, basis function is a set that help to conduct this approach were also known as basis smoothing. A fitted curve is formed from discrete data into functional objects that will be used for analyzation in future. Set of smooth functions can be formed through the linear combination of basis functions; $X(t) = \sum_{k=1}^{K} C_k \varphi_k(t)$ where C_k denotes basis coefficient while $\varphi_k(t)$ is basis function and K represents the size required for the basis.

The basis function used in this research is B-spline basis since it is involved with non-periodic data. A B-spline basis is defined over a non-decreasing sequence of knots $\{t_i\}$. This basis is piecewise polynomial segments jointed end-to-end at arguments values known as knot vector. It can be defined as $t = (t_0, t_1, \ldots, t_{n+k+1})$ which portray as an ordered set that consist of $n + k + 1$ elements where n is number of basis functions while k is order or degree of the B-spline polynomial. Then, the joint for each knot mapped into polynomial curve with $t_i \leq t \leq t_{i+1}$. Then, B-spline basis function can be expressed as follows for k =0, $B_{i,0}(t) = 1$ if $t_i \leq x < t_{i+1}$ while $B_{i,0}(t) = 0$ otherwise and basis functions for k >0 are defined as:

$$B_{i,k}(t) = \frac{x - t_i}{t_{i+k} - t_i} B_{i,k}(t) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(t) \quad (1)$$

where *x* indicates the specific points in which the basis functions are evaluated representing the weeks in data. In this study, predefined basis functions were used to smooth the data directly. Without explicitly estimating coefficients, the smoothed curve can be constructed by summing the predefined basis functions. The overall smoothing function f(x) is expressed as $f(x) = \sum_{i=1}^{n} B_i(x)$ where n represents parameter of basis function. Then, the B-spline basis functions, $B_i(x) = B_1(x), B_2(x), \ldots, B_n(x)$ are constructed based on the knots, t_i that are used by default number of knots which depends on the range of data and the number of basis functions through R programming software together with *k*=3 since cubic B-spline with degree of 3 was applied.

Besides, Fourier basis functions also was applied in this research which is suitable for present periodic data. This basis consists of cosines and sines with different frequencies. An interval of T where $(t_0, t_1, \ldots, t_j)\epsilon T$ was defined for periodic functions with the $\phi_n(t)$'s taking on the following form of $\phi_n(t) = 1$ if $n = 0$, $\phi_n(t) = \cos\left(\frac{2\pi nx}{T}\right)$ if $n > 0$ and even while $\phi_n(t) = \sin\left(\frac{2\pi nx}{T}\right)$ if $n > 0$ and odd. Then, periodic function with period T which is equal to 53 weeks can be expressed as Fourier series where $n = 1, \ldots, N$and N represent the index of frequency component or number of basis function. With n basis functions, the first N sine and cosine pairs are constructed since each pair counts as 2 functions and predefined as $\phi_0(x) = 1$, $\phi_n(x) = \cos\left(\frac{2\pi nx}{53}\right)$, $\phi_{n+N}(x) = \sin\left(\frac{2\pi nx}{53}\right)$ for $n = 1, 2, \ldots, N$ and $x = 1, 2, \ldots, 53$. Thus, the smoothing function $f(x)$ can be expressed as follows without any explicit coefficient estimation $f(x) = \phi_0(x) + \sum_{i=1}^{n}\left(\phi_n(x) + \phi_{n+N}(x)\right)$.

3.3. Error measurement

Error was calculated to determine and measure the accuracy between actual data and predicted data that have been obtained through the result of smooth curve generated where to distinguish which type of basis are the most suitable to plot smooth curve based on the data. Thus, type of error applied in this research are including Root Mean Squared Error (RMSE). Root Mean Squared Error (RMSE) was calculated by using following formula of $RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{n}}$ where n portrays the number of observations data, y_i represents the actual or observed values and $\hat{y}_i$ is predicted values [19] which can be obtained directly from the data point that was extracted through the smoothing curve which is generated by using the predefined basis functions with the assumption for all coefficient for both basis function is equal to 1. The less the error obtained shows the more accurate of the result.

Functional data analysis in R

Functional data analysis (FDA) was performed using R programming software through package of fda.usc which provides a set of tools for functional data analysis and modelling while codes such as create.bspline.basis() and create.fourier.basis() were used in this study to create B-spline and Fourier basis function before performing curve smoothing. Function for plotting functional data, such as plot() and line() were used in this study to visualize the functional data. These functions can be used to create plots of the raw data, as well as plots of fitted functions or curves.

Results and discussion

Preliminary Analysis

Basic statistical measure for weekly DHF cases in Malaysia over 12 years period are shown below.

Table 1: Summary Statistics of DHF cases over 12 years period

States/Federal Territory	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Johor	0.0000	0.0000	1.0000	1.9060	2.0000	24.0000
Kedah	0.0000	0.0000	0.0000	0.1840	0.0000	3.0000
Kelantan	0.0000	0.0000	0.0000	0.7060	1.0000	30.0000
Melaka	0.0000	0.0000	0.0000	0.8978	1.0000	26.0000
Negeri Sembilan	0.0000	0.0000	0.0000	0.4811	1.0000	8.0000
Pahang	0.0000	0.0000	0.0000	0.4041	0.0000	7.0000
Perak	0.0000	0.0000	0.0000	0.5094	1.0000	7.0000
Perlis	0.0000	0.0000	0.0000	0.0157	0.0000	1.0000
Pulau Pinang	0.0000	0.0000	0.0000	0.4371	1.0000	6.0000
Sabah	0.0000	0.0000	0.0000	0.3836	1.0000	5.0000
Sarawak	0.0000	0.0000	0.0000	0.5110	1.0000	10.0000
Selangor	0.0000	2.0000	5.0000	8.7800	11.0000	70.0000
Terengganu	0.0000	0.0000	0.0000	0.6525	1.0000	16.0000
WP Kuala Lumpur	0.0000	1.0000	2.0000	2.1490	3.0000	17.0000

Based on Table 1, the highest mean of weekly cases contributed by Selangor while WP Kuala Lumpur showed second highest mean value which is 2.1490. Furthermore, Johor recorded 1.9060 cases weekly over 12 years as its mean value and its value almost twice of Melaka states. Then, lowest cases occurred in Perlis which hold mean value of 0.0157 followed by second lowest mean value 0.1840 which was in Kedah. In addition, DHF weekly cases over periods of 12 years in Selangor is the highest maximum cases which is 70 cases when compared to other states while lowest maximum cases were in Perlis with 1 case. Meanwhile, there exist for all states and federal territory have no weekly cases since the minimum value for all is equal to zero. Thus, it shows that attention and precaution need to be directed given to Selangor because the value of cases is higher than other states for further actions.

Functional Data Analysis

The mean for weekly DHF cases in each state was calculated as an early step before plotting the data. Then, all the mean data is converted into functional data by using R programming with the code of Data2fd(). The two curves were generated by using B-spline and Fourier basis as to observe the pattern and trend of average DHF data cases for all states and federal territory in Malaysia.

The average DHF cases for B-spline curve in Johor as shown in Figure 1 begin to rise at the early week 1 to week 5 and slowly decrease until week 20. Then, it continues slowly increase until week 30 before the average cases reduce until the end of the year. Besides, Fourier smooth curve portrays almost same pattern with B-spline but difference at the beginning and the end of the year which is higher average cases when compared to B-spline. The difference is around 0.1 at the beginning and 0.5 at the end of the year. The curves of data in Johor depicts that average DHF cases are inconsistent throughout the year. Smooth curve of B-spline in Negeri Sembilan shows negative trend throughout the year where average cases decreased for the first 10 weeks and fluctuate until reach lowest cases at the end of the year. The smooth curve of Fourier basis is fluctuated by following the same trend with B-spline basis but lower by 0.2 average cases on the first week and greater by 0.3 average cases at the last week of the year. Then, smooth curve of Fourier basis also depicts almost stationary trend although it portrays similar pattern with B-spline basis curve. This shows that both type of basis produced different trend of curve of average cases along 53 weeks.

On the other hand, smooth curve for B-spline basis shows average cases in Kelantan fluctuated around 0.5 average cases for the first 20 weeks before drop and then rise rapidly started from week 22. The average cases reach the peak at week 32 before dies down and rise again at the end of year. This shown that average cases in Kelantan is high between week 25 and week 42. Based on the Fourier smooth curve, it presents fluctuated average cases similar to B-spline basis curve from the beginning to the end of the year but highest fluctuation cases occurred between week 28 and week 40 with greater value of average cases when compared to B-spline. This shown that average cases in Kelantan are fluctuated throughout the year but not involved high range of cases where it is only up to 1 average cases. It also shown that both basis presents almost same pattern and trend of the curve for Kelantan. Furthermore, trend of smooth curve for B-spline in Melaka depicts rising pattern at first 5 weeks and dies down slowly until week 10. Then, curve fluctuated slowly with positive trend until it reached at peak in week 44 before fall down rapidly until to the end of the year. It shown that average cases were increased and higher in second half of the year while kept reduce at last few weeks of the year. Besides, trend of smooth curve for Fourier basis shows almost the same trend as B-spline but it oscillates frequently and rapidly throughout the year with involved high range of average cases which is greater around 0.4 to 0.5 than B-spline. It also shown that curve for Fourier basis start and end with almost 0.2 average cases more than B-spline. Then, smooth curves that illustrate average DHF cases in Johor, Kelantan and Melaka were increased in the second half of the year over 12 years period.

Based on Figure 1, smooth curve of B-spline in Kedah starts high, drops quickly, then fluctuates before ending sharply at the end of the year. Average cases rise slowly from week 10 to 30, then spike between weeks 40 and 45 and showing an overall negative trend from week 1 to 53. The Fourier curve fluctuates throughout the year and peaking in the middle with higher average cases than the B-spline curve which indicating a stationary trend over 53 weeks. Thus, both type of basis demonstrated different trend of curve which will result of different analysis especially for the beginning and the end of the year.

Meanwhile, B-spline curve for Pahang depicts initially rises for 8 weeks, then falls until the end of the first quarter. It slowly rises until week 30, dips and peaks at week 45 before sharply declining by the end of the year. The Fourier curve follows a similar pattern in the first half of the year but fluctuates more and ending with higher average cases compared to B-spline. Generally, both type of basis function presented similar stationary trend with same pattern but only for the first half of the year.

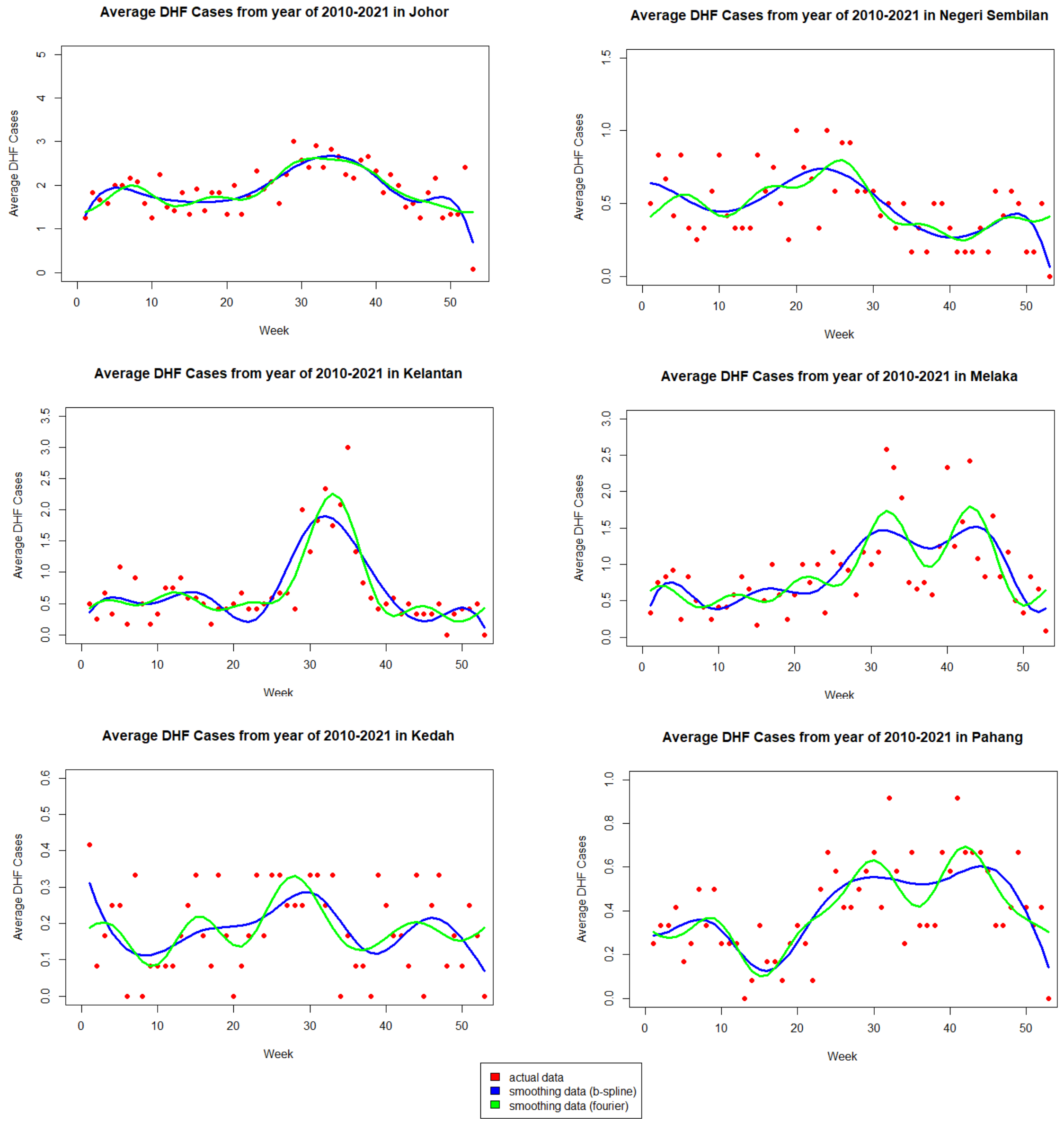

Figure 1 Smoothing curve for weekly average DHF cases over 12-year period in Johor, Negeri Sembilan, Kelantan, Melaka, Kedah and Pahang.

In addition, smooth curve of B-spline of Perak in Figure 2 declines for 5 weeks, then slowly rises for 5 weeks before declining until week 15. It fluctuates until week 50 before a rapid decline in the last 3 weeks. There are noticeable drops in average cases between weeks 10-20 and 40-50. The Fourier curve follows a similar pattern but starts lower and ends higher than the B-spline curve with slight differences in average cases. Then, each basis depicts almost similar pattern with different trend which also due to different average cases at the beginning and the end of the year. Besides, smooth curve of B-spline basis for Perlis depicts a consistent decrease in average cases starting from week 5. The first 10 weeks shows fluctuating average cases with the highest values compared to other weeks and continue to fluctuate with decreasing values until the end of the year. In the meantime, Perlis records mostly zero average cases in the second half of the year. The Fourier curve follows a similar trend but fluctuates more frequently with a negative trend. However, the average cases involved are small which is less than 0.1 that resulting in minor differences in changes. Generally, both of basis shows stationary trend of curve with similar pattern which gives almost same analysis towards DHF cases in Perlis.

Subsequently, smooth curve of B-spline for Pulau Pinang in Figure 2 portrays curve rises quickly, peaks in the first 5 weeks, then sharply drops until week 15. It slowly rises until week 30, then falls again until week 40. After that, it rises rapidly before another sharp drop at week 49. This shows higher average cases at the start and end of the year with a negative trend overall. The Fourier curve shows a similar pattern with rapid fluctuations but ends with a higher value compared to the B-spline curve. There are two peak points, occurring in the first 8 weeks and last 6 weeks of the year. The Fourier basis curve also shows stationary trend since first and last week recorded with same level of average cases. Besides, smooth curve of B-spline for Sabah shows a rapid increase in average cases for the first 10 weeks followed by a sharp decline until week 18. Then, it shows a positive trend with slow fluctuations until week 50 before sharply declining for the rest of the year. There is a negative trend over 53 weeks due to average cases on week 53 lower than week 1. The Fourier curve for Sabah follows a similar pattern but with rapid fluctuations throughout the year and higher average cases compared to B-spline. It peaks at week 9 with even higher average cases than the B-spline curve and shows a stationary trend throughout the year. Thus, both basis for both states illustrated similar pattern of curve with different trend that resulted with different analysis towards the average cases in Sabah.

The smooth curve of B-spline in Figure 2 shows Sarawak's average cases fluctuated rapidly during the first half of the year but these changes are relatively small. Despite higher average cases in the first half, there is a negative trend in the second half. Similarly, the Fourier curve follows a similar pattern in the first half but starts with lower average cases. In the second half, it fluctuates rapidly with higher average cases than B-spline but maintains a stationary trend over 53 weeks. Next, smooth curve of B-spline basis in Selangor depicts a negative trend throughout the year. Average cases fluctuate within a high range in the first half, then gradually decrease until week 40 before rising again until week 46. However, average cases drop rapidly after week 50. The Fourier curve follows a similar trend but fluctuates more frequently where starts with lower average cases than B-spline in the first week but ends higher in week 53. It also displays a stationary curve with similar pattern with B-spline. In general, B-spline and Fourier basis present curve with similar pattern but different trend that will lead to distinct analysis on Selangor's cases when compared to Sarawak.

Furthermore, smooth curve of B-spline in Figure 2 shows average cases in Terengganu rise rapidly for the first 3 weeks and decreased from highest peak at 1.3 average cases into 0.5 average cases in week 10 and kept decreased until week 23. Then, curve rise slowly until week 38 and keep fluctuate before rapidly fall down at week 50 until to the end of the year. This curve depicts stationary trend from first to the last 53 weeks which also shows decreasing of average cases throughout the year. In the meantime, smooth curve of Fourier illustrated similar pattern with B-spline but fluctuated slowly for the first half of the year and then fluctuated rapidly for the second half of the year. The average cases for Fourier basis are lower at beginning of the year and higher at the end of the year when compared to B-spline which is different around 0.1 to 0.6 average cases. The smooth curve for Fourier basis shows stationary trend although it almost same pattern with B-spline curve. Thus, it shown that both basis functions demonstrated different analysis due to different trend of curves. Last but not least, smooth curve of B-spline and Fourier for WP Kuala Lumpur presents quite same pattern of data with Selangor but it is involved small range of average cases over 12-years period. There are different on the highest point reached at the first 5 weeks for the B-spline curve when compared to Selangor. It shown that average cases decrease with fluctuation along 53 weeks and leads to negative trend. Then, Fourier smooth curve presented similar pattern as B-spline curve but only different at first 5 weeks where the average cases increased slowly when compared to B-spline. Besides, curve dies down early at week 45 before it ends with highest average cases when compared to B-spline curve. The trend of this Fourier basis curve depicts the stationary trend of average cases along 53 weeks. Overall, smooth curves of Fourier basis in majority of the states shows stationary trend while B-spline curve depicts negative trend.

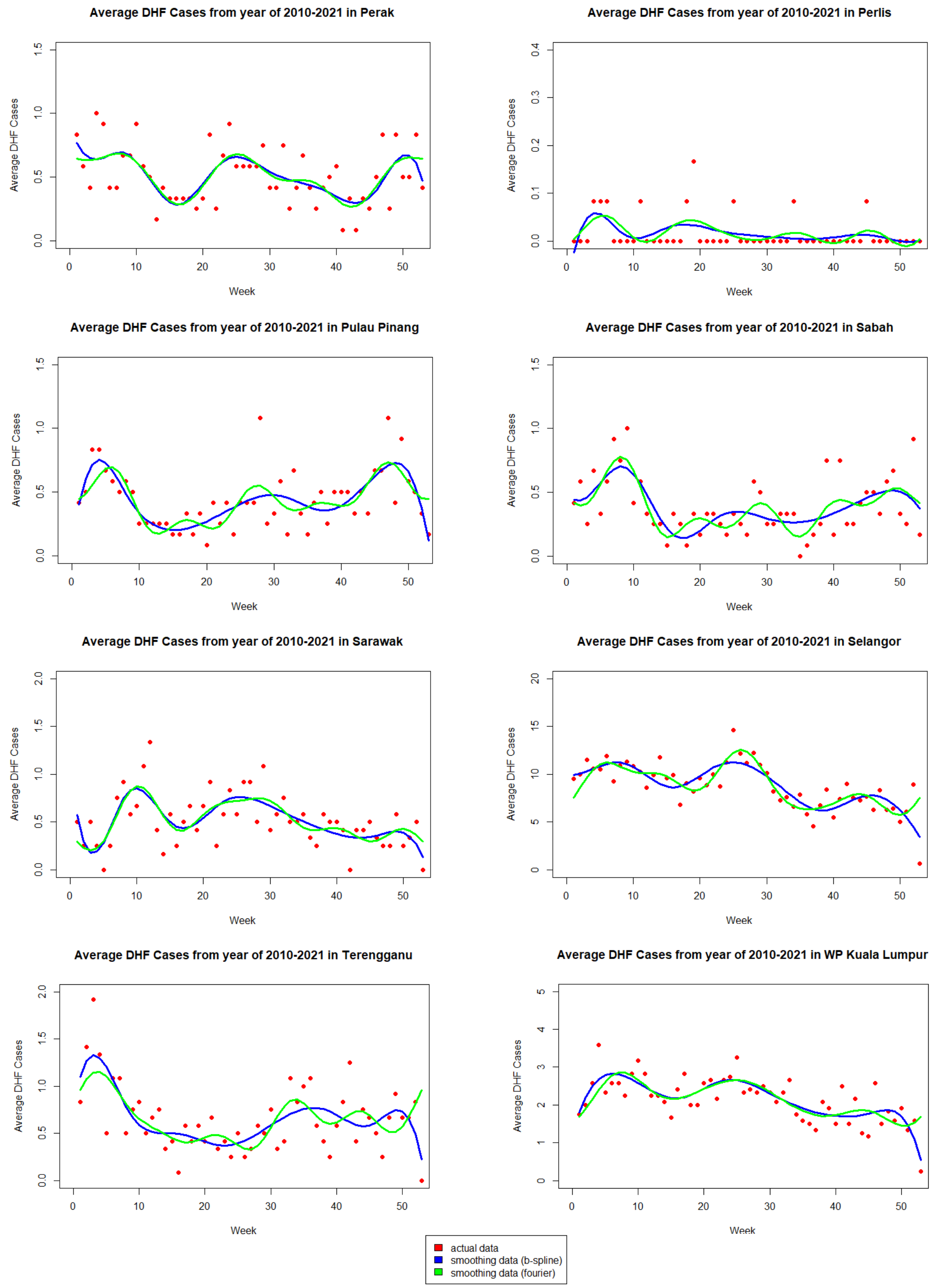

Figure 2 Smoothing curve for weekly average DHF cases over 12-year period in Perak, Perlis, Pulau Pinang, Sabah, Sarawak, Selangor, Terengganu and WP Kuala Lumpur.

The smooth curves in Figure 3 are clearly shown that Selangor have negative trend while other states depict almost stationary trend throughout 53 weeks. Besides, it also depicts that the average cases for Selangor is decreasing with rapid fluctuation pattern throughout the year over 12 years period. In the meantime, Selangor record highest average DHF cases which is up to 12 for Fourier curve and up to 11 for B-spline curve when compared to other states while the lowest average DHF cases contributed by Perlis. Average DHF cases for other states were only involved within range up to 3 average cases. In addition, smoothing plot above also shows the top 3 highest average DHF cases after Selangor which is WP Kuala Lumpur, Johor and Kelantan.

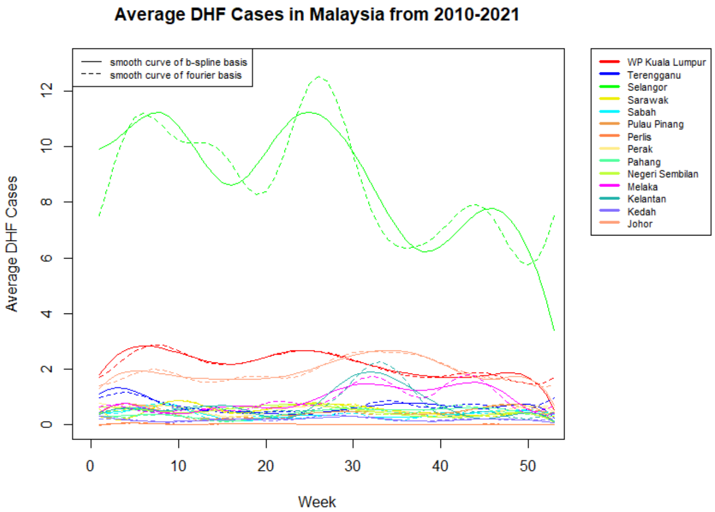

Figure 3 Smoothing curve of weekly average DHF cases in Malaysia over 12-year period.

Based on Table 2, calculated RMSE shows errors for each state and federal territory for both basis in order to make comparison in selecting suitable type of basis should be used. The B-spline smooth curve is suitable to be applied for data of Johor as to observe the trend and pattern of average DHF cases followed by Negeri Sembilan, Perak, Pulau Pinang, Sarawak, Selangor, Terengganu and WP Kuala Lumpur since it is involved small error between the actual data. On the other hand, other states such as Kedah, Kelantan, Melaka, Pahang, Perlis and Sabah are suitable to apply Fourier smooth curve due to small error produced when compared to B-spline smooth curve. From this RMSE, highest error involved Selangor for both basis since it is involved high range of average cases while Perlis is the smallest error among others since it is only involved with small range of average cases.

Table 2: RMSE comparison between B-spline and Fourier basis for 13 states and 1 federal territory.

States/Federal Territory	B-spline	Fourier
Johor	0.3397	0.3622
Kedah	0.1035	0.1027
Kelantan	0.3386	0.2841
Melaka	0.4161	0.3825
Negeri Sembilan	0.1832	0.1926
Pahang	0.1453	0.1396
Perak	0.1799	0.1808
Perlis	0.0331	0.0324
Pulau Pinang	0.1569	0.1655
Sabah	0.1760	0.1636
Sarawak	0.2059	0.2072
Selangor	1.3676	1.4278
Terengganu	0.2520	0.2762
WP Kuala Lumpur	0.3733	0.4306

Then, RMSE depicts that majority of the states and federal territory are suitable to use B-spline basis than Fourier basis for curve smoothing since small error obtained which is also represent more accurate result. Thus, it can be concluded that B-spline basis is the most suitable basis could be used in performing smooth curve as to observe trend and pattern of average DHF cases for most of the states which have non-periodic pattern of data. Furthermore, suitable number of basis functions is necessary to smooth the curve since it functions in determining the flexibility of curve. The selection number of basis functions used also was analyzed by executing the comparison of RMSE through 6 different number starting from 7 to 12 towards 3 out of 13 states and 1 federal territory which is Johor, Kedah and Selangor as shown in Table 3. These states with 2 different type of basis show that uses of larger number of basis function will result of smaller values of error. Besides, Fourier basis depict that increases of each 2 number of basis function will give the effect to the reducing of error. This proved that accuracy of the smooth basis is high when larger number of basis function has been used.

Table 3: RMSE through 6 different number of basis function for Johor, Kedah and Selangor.

States / Number of basis function	Johor		Kedah		Selangor	
	B-spline	Fourier	B-spline	Fourier	B-spline	Fourier
7	0.3683	0.3736	0.1060	0.1090	1.4121	1.5330
8	0.3530	0.3686	0.1061	0.1033	1.5104	1.4784
9	0.3449	0.3686	0.1044	0.1033	1.2683	1.4784
10	0.3397	0.3622	0.1035	0.1027	1.3676	1.4278
11	0.3377	0.3622	0.1007	0.1027	1.1740	1.4278
12	0.3309	0.3514	0.1003	0.0988	1.1774	1.4265

In the meantime, smoothness of the curve also needs to be considered since it controls how much the curve can deviate from the data points. Based on Figure 4 and Figure 5, the best smooth curve for B-spline is occurred at maximum number of basis function of 10 where there is no exist of any sharp curve at any point while for Fourier basis in Johor and Kedah is 11. Then, Selangor was selected in the analysis since it is involved highest range of average DHF cases compared to others. The most suitable number of basis function to be used for data of Selangor as shown in Figure 6 is 10 or 11. In conclusion, after take consideration to the value of error and the smoothness of curve between three states that have difference range of average DHF cases, it gives the result of 10 for the suitable number of basis function to be used for all states in Malaysia.

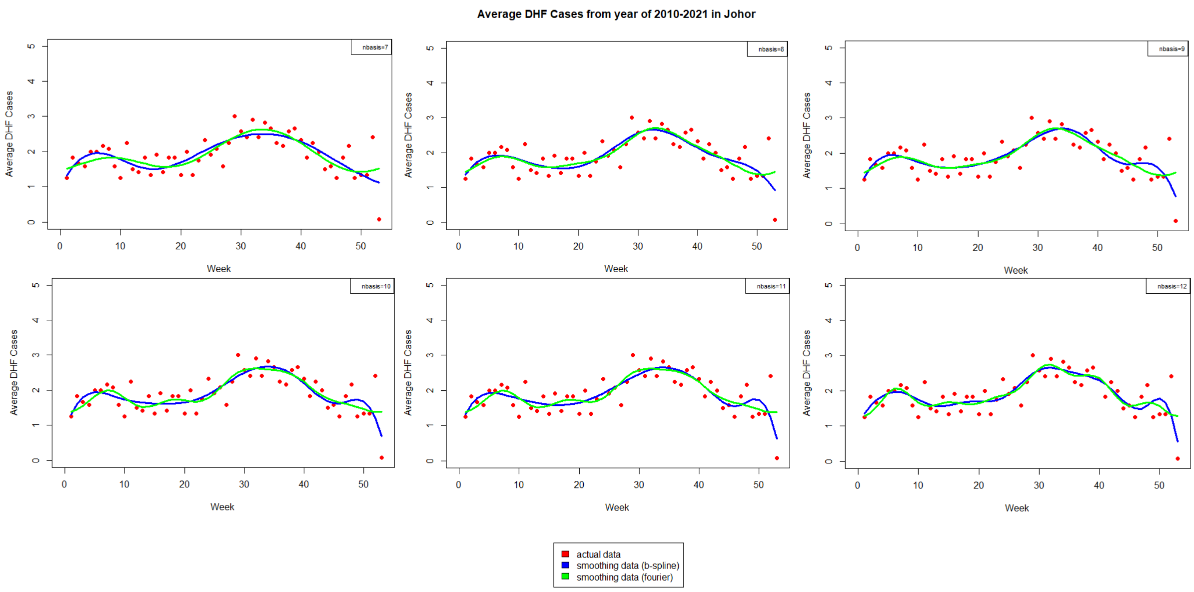

Figure 4 Smooth curve of B-spline and Fourier basis functions for different number of basis functions for state of Johor.

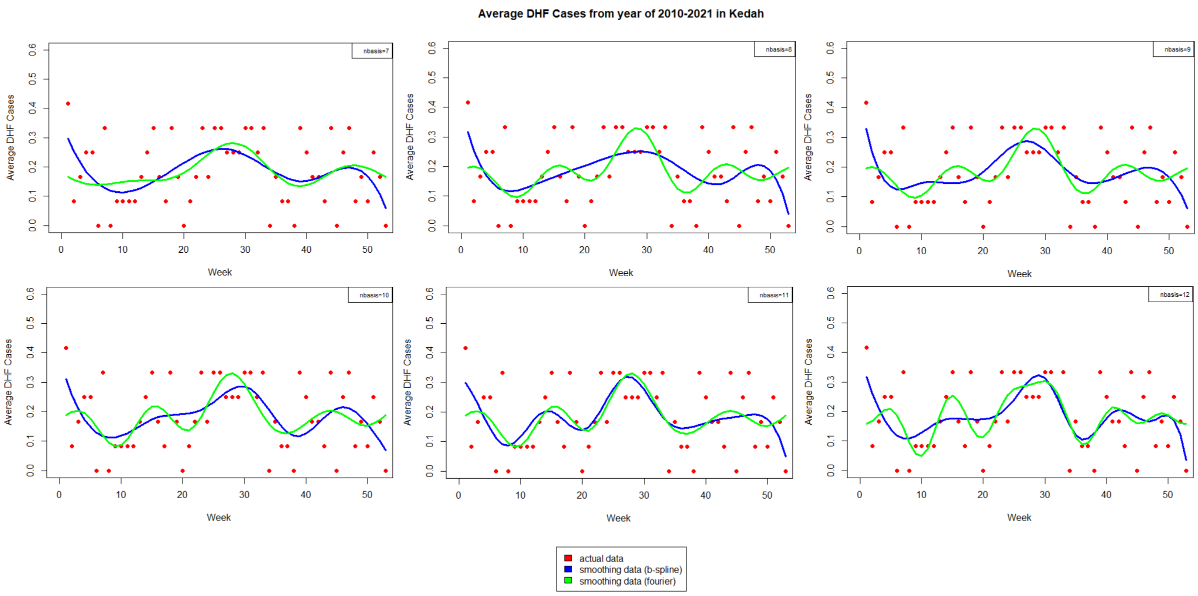

Figure 5 Smooth curve of B-spline and Fourier basis functions for different number of basis functions for state of Kedah.

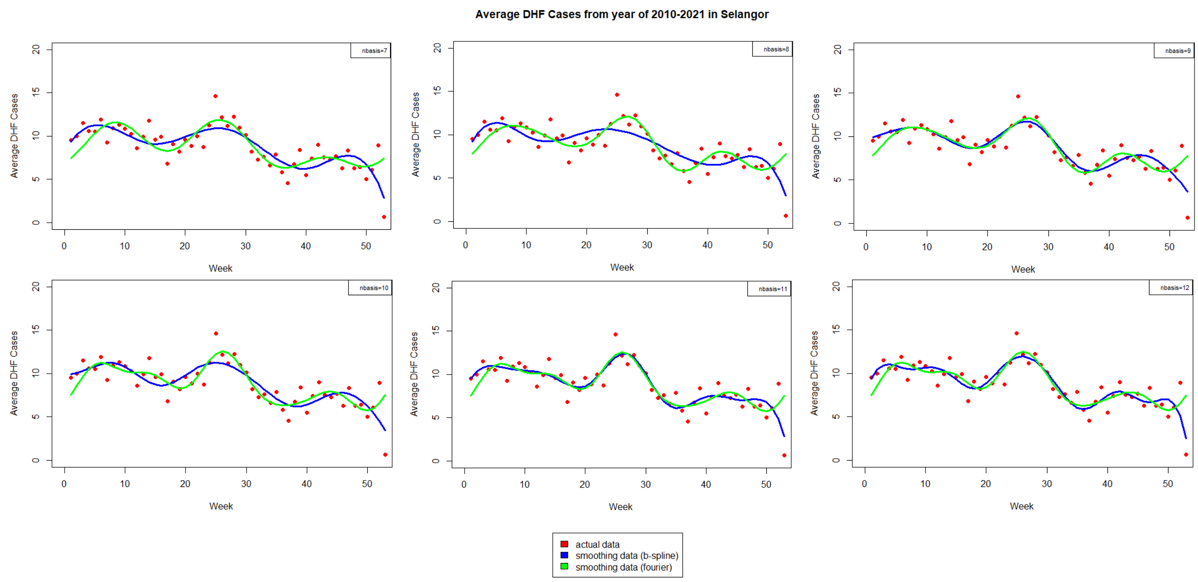

Figure 6 Smooth curve of B-spline and Fourier basis functions for different number of basis functions for state of Selangor.

Conclusion

This research successfully identified patterns and trends in DHF cases data from 13 states and 1 federal territory in Malaysia over a 12-year period. FDA techniques used in this study is not only provided the visualization of DHF cases data but also offer a suitable basis with appropriate number of basis function. The analysis of this study depicts the B-spline basis function is the most suitable basis function to be used for these datasets in plotting graph and smooth curve since it is very helpful to observe pattern and trend of DHF cases. The findings of this research through applying the FDA were managed to provide better understanding on the underlying pattern of the data without noise found in the raw data.

Acknowledgement

The authors gratefully acknowledge the Disease Control Division in Ministry of Health Malaysia and Dr. Muhammad Fauzee Hamdan from Universiti Teknologi Malaysia (UTM) for providing the weekly DHF data and information for this study.

References

[1] Bernama. 2023. Dengue cases in 2022 rose by 150%, death toll up by 180% from 2021: Health DG. The Sun. Retrieved October 26, 2023, from https://thesun.my/local/dengue-cases-in-2022-rose-by-150-death-toll-up-by-180-from-2021-health-dg-IA10452497

[2] Mayo Clinic. 2022. Dengue fever - Symptoms and causes. Retrieved October 26, 2023, from https://www.mayoclinic.org/diseases-conditions/dengue-fever/symptoms-causes/syc-20353078

[3] Wang, J., Chiou, J., & Mueller, H. 2015. Review of Functional Data Analysis. ArXiv. DOI: https://doi.org/10.48550/arXiv.1507.05135

[4] Mazelan, N. A., & Suhaila, J. (2023). Exploring rainfall variabilities using statistical functional data analysis. In IOP Conference Series: Earth and Environmental Science (Vol. 1167). Institute of Physics. DOI: https://doi.org/10.1088/1755-1315/1167/1/012007

[5] Boschi, T., Di Iorio, J., Testa, L., Cremona, M. A., & Chiaromonte, F. (2021). Functional data analysis characterizes the shapes of the first COVID-19 epidemic wave in Italy. Scientific Reports, 11(1). DOI: https://doi.org/10.1038/s41598-021-95866-y

[6] Matabuena, M., Vidal, J. C., Hayes, P. R., Saavedra-Garcia, M., & Trillo, F. H. (2019). Application of Functional Data Analysis for the Prediction of Maximum Heart Rate. IEEE Access, 7, 121841–121852. DOI: https://doi.org/10.1109/ACCESS.2019.2938466

[7] Tian, T. S. (2010). Functional data analysis in brain imaging studies. Frontiers in psychology, 1, 35. DOI: https://doi.org/10.3389/fpsyg.2010.00035

[8] Capezza, C., Centofanti, F., Lepore, A., & Palumbo, B. (2021). A functional data analysis approach for the monitoring of ship CO2 emissions. Gestão & Produção, 28(3), e152. DOI: https://doi.org/10.1590/1806-9649-2021v28e152

[9] Kutrolli. G., & Benth F. E. (2019). An Application of Functional Data Analysis to Forecast Weather Variables. DOI: https://doi.org/10.2139/ssrn.3766459

[10] Shaadan, N., Deni, S. & Jemain, A. 2015. Application of Functional Data Analysis for the Treatment of Missing Air Quality Data. Journal of Sains Malaysiana. 44(10). 1531-1540. DOI: http://doi.org/10.17576/jsm-2015-4410-19

[11] Ismail, W. N. & Jamaludin, S. S. 2015. Smoothing Wind and Rainfall Data through Functional Data Analysis Technique. Journal of Technology. 74. 105-112. DOI: https://doi.org/10.11113/jt.v74.3011.

[12] Oshinubi, K., Ibrahim, F., Rachdi, M., & Demongeot, J. 2022. Functional data analysis of COVID-19 dynamics in French departments before and after vaccination. AIMS Mathematics, 7(4), 5347-5385. DOI: https://doi.org/10.3934/math.2022298

[13] Bett, B., Grace, D., Lee, H. S., Lindahl, J., Nguyen-Viet, H., Phuc, P. D., … Nam, V. S. 2019. Spatiotemporal analysis of historical records (2001–2012) on dengue fever in Vietnam and development of a statistical model for forecasting risk. PLoS ONE, 14(11). DOI: https://doi.org/10.1371/journal.pone.0224353

[14] Guo, C., Zhou, Z., Wen, Z., Liu, Y., Zeng, C., Xiao, D., … Yang, G. 2017. Global epidemiology of dengue outbreaks in 1990–2015: A systematic review and meta-analysis. Frontiers in Cellular and Infection Microbiology, 7(JUL). DOI: https://doi.org/10.3389/fcimb.2017.00317

[15] Majid, N. A., Nazi, N. M., & Mohamed, A. F. 2019. Distribution and spatial pattern analysis on dengue cases in Seremban District, Negeri Sembilan, Malaysia. Sustainability (Switzerland), 11(13). DOI: https://doi.org/10.3390/su11133572

[16] Chengathir, S. M., Bhuvaneswari, T., Maruthupandi, J., & Naga, P. R. 2022. Prediction of dengue using data mining classification algorithms. International Journal of Health Sciences. DOI: https://doi.org/10.53730/ijhs.v6ns1.7907

[17] Mohsin, M. F. M., Bakar, A. A., Hamdan, A. R., Sahani, M., & Ali, Z. M. 2023. Dengue Outbreak Detection Model Using Artificial Immune System: A Malaysian Case Study. Journal of Information and Communication Technology, 22(3), 399–419. DOI: https://doi.org/10.32890/jict2023.22.3.4

[18] Husin, N. A. & Salim, N. 2008. A comparative study for back propagation neural network and nonlinear regression models for predicting dengue outbreak. Journal of Information Technology, 20 (4). pp. 97-112. ISSN 0128-3790.

[19] Hodson, T. O. 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development. Copernicus GmbH. DOI: https://doi.org/10.5194/gmd-15-5481-2022